Supplementary Materials

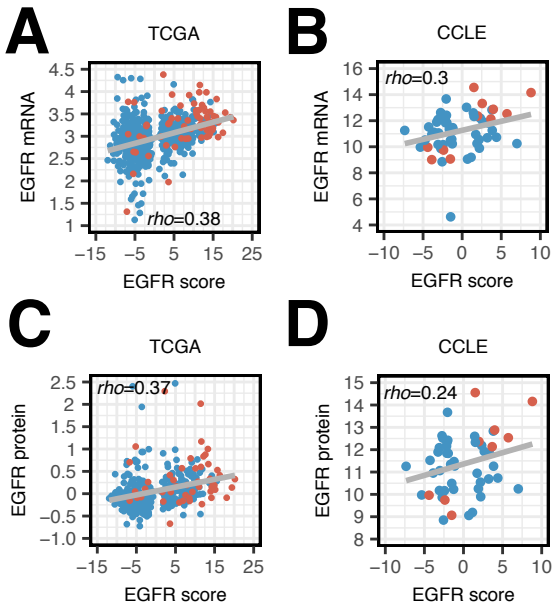
An EGFR signature predicts cell line and patient sensitivity to multiple tyrosine kinase inhibitors

Chao Cheng[1,2,3*], Yanding Zhao[4], Evelien Schaafsma[4], Yi-Lan Weng[5], Christopher Amos[1,2,3]

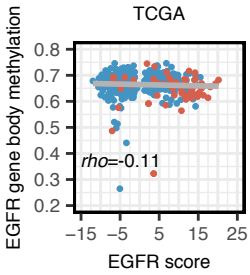
Table of contents

Supplementary Figures
Suppl. Figure 1. Correlation between EGFR scores and EGFR mRNA abundance.
Suppl. Figure 2. Correlation between EGFR scores and the average methylation level of CpG sites at the EGFR gene body region.
Suppl. Figure 3. Boxplot depicting EGFR score differences between Erlotinib sensitive and resistant lung cancer cell lines.

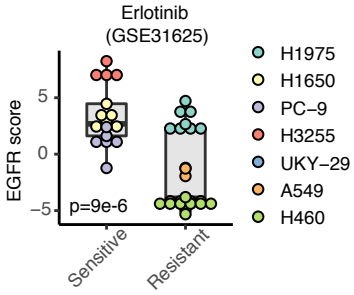
Supplementary Tables
Suppl. Table 1: Overview of utilized datasets.
Suppl. Table 2: Weights of genes in the EGFR signature.
Suppl. Table 3: Pathway enrichment analysis.
Suppl. Table 4: Identifiers of samples used in this study.



Suppl. Figure 1. A-B, Correlation between our EGFR score and EGFR mRNA abundance in TCGA lung adenocarcinoma tumor samples **(A)** and CCLE lung cancer cell lines **(B)**. **C-D**, Correlation between EGFR scores and EGFR protein abundance in TCGA lung adenocarcinoma tumor samples **(C)** and CCLE lung cancer cell lines **(D)**. In all scatterplots, ρ indicates Spearman correlation coefficient.



Suppl. Figure 2. Correlation between EGFR scores and the average methylation level of CpG sites at the EGFR gene body region. ρ indicates Spearman correlation coefficient.



Suppl. Figure 3. Boxplot depicting EGFR score differences between Erlotinib sensitive and resistant lung cancer cell lines. The p-value was calculated by the one-tailed Wilcoxon rank-sum test.

Suppl. Table 1: Overview of utilized datasets. For each dataset, its GEO/dataset ID, platform, author, PMID, survival information type and number of samples are provided.

GEO/dataset ID	Platform	Platform annotation	Author	PMID	EGFR mutant	WT	Samples	Treatment	Tissue origin
TCGA	Illumina Hiseq 2000	RNA-seq	NA	25079552	72	443	515	No	NSCLC tumor tissue
GSE13213	GPL6480	two channel	Tomida	19414676	45	72	117	No	NSCLC tumor tissue
GSE31210	GPL570	one channel	Okayama	22080568	127	119	246	No	NSCLC tumor tissue
GSE11969	GPL7015	two channel	Takeuchi	16549822	34	115	149	Yes	NSCLC tumor tissue
GSE31852	GPL6244	one channel	Saintigny	22586319	20	104	124	Yes	NSCLC tumor tissue
GSE26939	GPL9053	two channel	Wilkerson	22590557	11	70	81	No	NSCLC tumor tissue
GSE32989	GPL13376	one channel	Byers	23091115	9	32	41	Yes	NSCLC cell line
GSE31625	GPL96	one channel	Balko	17096850	NA	NA	42	Yes	NSCLC cell line
CCLE	GPL15308	one channel	NA	22460905	6	45	51	Yes	NSCLC cell line
GDSC	GPL13367	one channel	NA	23180760	5	35	40	Yes	NSCLC cell line
GSE109211	GPL13938	one channel	Pinyol	30108162	NA	NA	67	Yes	HCC tumor tissue

Suppl. Table 3: Pathway enrichment analysis. For each enrichment term of genes in EGFR signature, its pathway term, geneset size, number of overlapped genes, enrichment ratio, P-value and FDR are provided.

Term	Geneset Size	Count	Enrichment ratio	PValue	FDR
KEGG_ASTHMA	28	11	9.67	5.30E-09	5.35E-06
KEGG_INTESTINAL_IMMUNE_NETWORK_FOR_IGA_PRODUCTION	46	13	6.96	2.19E-08	1.11E-05
KEGG_ALLOGRAFT_REJECTION	35	11	7.74	7.94E-08	2.67E-05
KEGG_GRAFT_VERSUS_HOST_DISEASE	37	11	7.32	1.51E-07	3.81E-05
KEGG_TYPE_I_DIABETES_MELLITUS	41	11	6.60	4.81E-07	9.70E-05
KEGG_LEISHMANIA_INFECTION	70	14	4.92	6.95E-07	1.17E-04
KEGG_VIRAL_MYOCARDITIS	68	13	4.71	2.96E-06	4.26E-04
KEGG_AUTOIMMUNE_THYROID_DISEASE	50	11	5.42	4.07E-06	5.13E-04
KEGG_CELL_ADHESION_MOLECULES_CAMS	131	18	3.38	5.98E-06	6.04E-04
KEGG_ANTIGEN_PROCESSING_AND_PRESENTATION	81	13	3.95	2.18E-05	1.83E-03
REACTOME_TRANSLOCATION_OF_ZAP_70_TO_IMMUNOLOGICAL_SYNAPSE	13	6	11.36	5.93E-06	6.04E-04
REACTOME_PHOSPHORYLATION_OF_CD3_AND_TCR_ZETA_CHAINS	15	6	9.85	1.61E-05	1.48E-03
REACTOME_PD1_SIGNALING	17	6	8.69	3.72E-05	2.88E-03
REACTOME_GENERATION_OF_SECOND_MESSENGER_MOLECULES	26	7	6.63	5.95E-05	4.29E-03
REACTOME_MHC_CLASS_II_ANTIGEN_PRESENTATION	90	12	3.28	2.78E-04	1.87E-02